International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 5, Issue 5, Oct 2015, 35-42

CLASSIFICATION OF MAMMOGRAPHIC BREAST DENSITY USING GREY LEVEL DIFFERENCE STATISTICS AND FOURIER POWER SPECTRUM FEATURES

VIPUL SHARMA[1] & LOKESH JAIN[2]

[1]Assistant Professor, Computer Science & Engineering, DAV University, Jalandhar, Punjab, India

[2]Assistant Professor, School of Electrical Engineering & Information Technology,
Punjab Agricultural University, Ludhiana, Punjab, India

ABSTRACT

Breast tissue density has been shown to be related to the risk of development of breast cancer, since dense breast tissue can hide lesions, causing the disease to be detected at later stages. Thus there is a need for the development of efficient techniques for the classification of breast density. In the proposed work, breast density (Fatty and Dense) is used as a pattern for classification. For carrying out the experiments mini-MIAS database has been used. This database contains images from screening mammography and has been widely used in the recent research. Texture features based on Grey Level Difference Statistics and Fourier Power Spectrum have been used for representing the texture pattern of fatty and dense mammograms. These features are then subsequently fed to the SVM classifier to classify fatty and dense mammograms. The results show that the extracted features perform very well with Polynomial Kernel of Support Vector Machine (SVM) classifier giving an accuracy of 97.25%. The experimental results encourage the use of proposed method for the classification of breast density.

KEYWORDS: Breast Density, Cancer, Texture Features, Classification, CAD

INTRODUCTION

In spite of good advancements for diagnosis and treatment, cancer is still a big threat to the society. Breast cancer is the most frequently diagnosed malignancy found in women [1-3]. According to the facts from the International Agency for Research on cancer, breast cancer comprises of 22.9% of invasive cancers affecting women. The data from the Population Based Cancer Registry (PBCR) also reveals that around 25-32% women living in metro cities of India are affected by this abnormality [4]. In Punjab state, around 4000 cases of various types of cancer have been reported during the last six months and most of the cases are of breast cancer. The worst affected are districts falling under the Malwa region of the Punjab. With widespread acceptance of mammography as a screening tool, there is a need to process efficiently such images using techniques of computer vision. Previous studies have shown that the sensitivity in detecting a breast cancer decreases due to increase in the breast density, as high density makes it difficult for the radiologists to see an abnormality which leads to false negative results [5-7]. Radiologists primarily estimate breast density by visual judgment of mammogram which is highly subjective. Automatic breast density classification methods attempt to mimic such visual judgment and classify images on the basis of underlying texture characteristics [8]. With widespread acceptance of mammography as a screening tool, there is a need to process images efficiently using techniques of computer vision.

This paper presents a scheme for the classification of fatty and dense mammograms based on texture features extracted using Grey level Difference Statistics (GLDS) and Fourier Power Spectrum (FPS) Model. The extracted features are finally fed to the Support Vector Machine (SVM) classifier for classification. The proposed methodology consists of various steps like Region of Interest Extraction, Texture Feature Extraction and Classification.

The remaining paper is organized as follows: Section II describes the related work done in the area of breast density classification. Section III describes the material and methods. In section IV, results and discussions are presented. Finally conclusions are given in section V.

LITERATURE REVIEW

In the context of mammography and breast cancer, some works have explored the use of Computer Aided Diagnosis (CAD) system.Oliver et al. proposed a CAD system for classification of breast density using morphological and texture features [8]. A set of 322 images from mini-MIAS database and 831 images from Digital Database for Screening Mammography (DDSM) database were used for evaluating the performance of the proposed system. For classification a Decision Tree classifier, Bayesian classifier and k-Nearest Neighbor were used. Bovis et al. proposed an approach for the classification of mammograms on the basis of breast density [9]. A total of 377 mammograms from DDSM were selected to evaluate the performance. The authors investigated the use of Spatial Grey Level Dependency (SGLD) matrices, Fourier Power Spectrum (FPS), Law's Texture Energy Measure, Discrete Wavelet Transform (DWT) based features for classifying mammograms. Subashini et al. proposed an automatic approach for assessing the breast tissue density [10]. The mini-MIAS database was used to evaluate the performance. Various statistical features were extracted from the ROI to represent the texture. Support Vector Machine (SVM) was used as a classifier to classify the images. Tzikopoulos et al. presented an approach for automatic segmentation of breast and classification scheme for breast tissue density estimation [11].The proposed algorithm was tested on mini-MIAS database. From each image first-order statistical features and fractal features were extracted. Classification was done by SVM. Ibrahim et al. proposed an approach for the classification of breast masses [12]. The authors have extracted sixty one features for the classification purpose based on the proposed visual method. For classification they have used k-NN and Support Vector Machine.

From the literature studied it has been found that one of the challenging aspects of CAD systems is to extract features from the images to represent efficiently their diagnostic and visual information content. There are many other issues to be considered in the design of a Computer Aided Diagnosis system that includes Region of Interest (ROI) extraction, feature extraction, selection of optimal features from the extracted features and classification.

Although a lot of work has been done in the area of breast density classification, but still it is the subject of great importance and relevance due to increasing prevalence of breast cancer across the globe.

MATERIALS AND METHODS

- **Mammogram Database Used**

For carrying out the proposed work mini-MIAS [13] database has been used. This standard database contains 322 images in Medio Lateral-Oblique (MLO) view. The original MIAS database (digitized at 50 micron pixel edge) has been reduced to 200 micron pixel edge and clipped/padded so that every image is 1,024 × 1,024 pixels. The images in database have ground truth provided by the experienced radiologists that includes location of abnormality, radius of circle enclosing the abnormality, character of background tissue and severity of abnormality. The images in this database are classified in

three categories based on their density as fatty, fatty-glandular and dense glandular.In this study, all the Fatty-Glandular and Dense-Glandular mammograms are treated as one group of dense mammograms giving a two class classification problem (Fatty, Dense). Table 1 shows the Division of Breast Density Classes in mini-MIAS Database.

Table 1: Division of Breast Density Classes in Mini-MIAS Database

Class	Character of Background Tissue	No. of Images
I	Fatty (F)	106
II	Fatty-glandular (G)	104
III	Dense-glandular (D)	112
Total Images		**322**

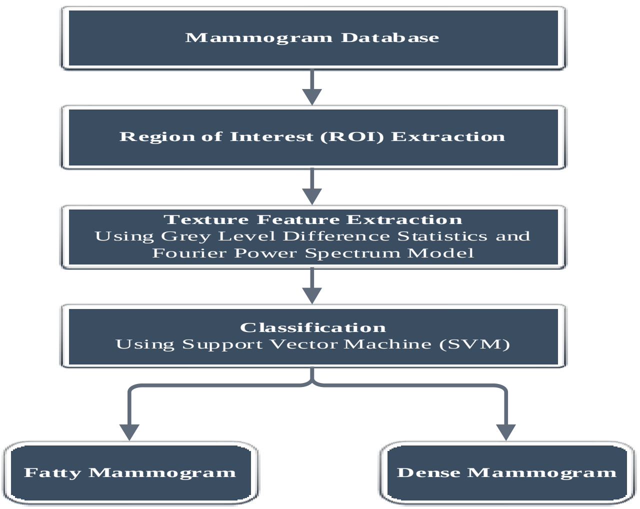

Figure 1: Block Diagram of Proposed Methodology

- **Region of Interest and Texture Feature Extraction**

The block diagram of proposed methodology is given in figure 1. For carrying out the experiments Region of Interest of size 200×200 pixels were manually cropped from the centre of breast tissue immediately behind the nipple in such a way that ROI contains tissue pattern only excluding the pectoral muscle and background area [14-16]. The fatty mammogram visually differs from dense mammogram in terms of tonal variations (intensity-based like contrast, brightness) as high density looks brighter in the mammography.In the proposed work texture features are extracted using Grey Level Difference Statistics Model and Fourier Power Spectrum Model.

- **Grey Level Difference Statistics Features**

The GLDS algorithm uses first-order statistics of local property values based on the absolute differences between pairs of gray levels or of average gray levels to extract the following 5 texture measures: Homogeneity, Contrast, Mean, Energy and Entropy [17]. These features are based on the absolute difference between pairs of gray levels separated at distance $\delta = (\Delta x, \Delta y)$. For a given displacement $\delta = (\Delta x, \Delta y)$, the difference image $f_\delta(x, y)$ is defined as:

$$f_\delta(x,y) = |f(x,y) - f(x + \Delta x, y + \Delta y)| \quad (1)$$

and $p_\delta$is the probability density (gray-level histogram) of $f_\delta(x,y)$ for m gray levels. Various texture features extracted from $p_\delta$are:

- **Homogeneity (HOMG)**: It is a measure of similarity in grey level intensities.

$$HOMG = \frac{\sum p_\delta(i)}{1+i} \tag{2}$$

- **Contrast (CNTG):** It is a measure of grey level intensity difference between neighboring pixels.

$$CNTG = \sum i^2 p_\delta(i) \tag{3}$$

- **Mean (MENG):** It is the average value of the grey level intensities within a given area.

$$MENG = \frac{1}{m}\sum i\, p_\delta(i) \tag{4}$$

- **Energy (ENGG):** It represents amplitude of grey level values.

$$ENGG = \sum p_\delta(i)^2 \tag{5}$$

- **Entropy (ENTG):** It measures the randomness in grey level intensities within the given area.

$$ENTG = -\sum p_\delta(i) \log p_\delta(i) \tag{6}$$

- **Fourier Power Spectrum Features**

This texture model contains the information on the texture orientation, grain size, and texture contrast of the image. The Discrete Fourier Transform (DFT) approach is used here for texture quantification because repetitive global patterns are difficult to describe with spatial techniques but relatively easy to represent with peaks in the spectrum [18]. The Radial sum and the Angular sum of the DFT were computed to describe texture. FPS features are computed from the power spectrum in the frequency domain.

$$|F(u,v)|^2 = F(u,v)F^*(u,v) \tag{7}$$

where, $F(u,v)$ is the Fourier transform of the image and $F^*(u,v)$ is the complex conjugate of Fourier transform of the image.

Spectral features are expressed in polar coordinates to yield a function $S(r,\theta)$. For each direction θ, $S(r,\theta)$ can be expressed as $S_\theta(r)$ and similarly for each frequency r, $S(r,\theta)$ can be expressed as $S_r(\theta)$. Analyzing $S_\theta(r)$ for a fixed value of θ gives the behaviour of spectrum along a radial direction from the origin and is called wedge analysis whereas analyzing $S_r(\theta)$ for a fixed value of r gives the behavior of spectrum along a circle centered on the origin and is called ring analysis. A global interpretation is obtained by summing over discrete variables:

$$S_\theta = \sum_{\theta=0}^{\pi} S_\theta(r) \tag{8}$$

And

$$S_r = \sum_{r=1}^{R_0} S_r(\theta) \tag{9}$$

Where, R_0 is the radius of circle centered at origin.

In this texture model, two features: S_r and S_θ are calculated and these are measure of the orientation of the texture.

- **Classification of Breast Density**

For classifying the mammograms into fatty and dense classes, Support Vector Machine (SVM) [19] classifier has

been used. SVM classifier is widely used in the recent research [20-24]. SVM guides the construction of classifier with good degree of generalization i.e. it has capability of predicting the class of sample that was not used in the learning process. For binary classification, SVM can be described as follows: Given two classes and set of points that belongs to these classes, the SVM classifier determines the hyper-plane in the projected feature space that separates the points in order to place the highest number of points of the same class on the same side, while maximizing the distance of each class to that hyper-plane. In some cases, the dataset cannot be precisely separated by a hyper-plane, so a kernel function is used. In this work, three kernel functions have been used namely, Polynomial Kernel, RBF Kernel and Pearson VII function-based universal kernel for evaluating the performance. For classifying the mammograms in two classes WEKA data mining tool has been used.

RESULTS & DISCUSSIONS

The proposed work for classifying the mammograms into two categories based on their texture has been done on 322 images of mini-MIAS dataset. The performance of the classifier is evaluated in terms of Sensitivity, Specificity, Accuracy and classification results are shown in Table 2.

Table 2: Classification Performance using Different Kernel Functions

Kernel Used: Polynomial								
Test Result	Actual		Sensitivity (%)	Specificity (%)	Accuracy (%)	Positive Predictive Power	Negative Predictive Power	Misclassification Rate
	Dense (Positive)	Fatty (Negative)						
Dense (Positive)	75 (TP)	3 (FP)	100	90.6	97.25	0.9625	1	0.0275
Fatty (Negative)	0 (FN)	29 (TN)						
Kernel Used: RBF Kernel								
Test Result	Actual		Sensitivity (%)	Specificity (%)	Accuracy (%)	Positive Predictive Power	Negative Predictive Power	Misclassification Rate
	Dense (Positive)	Fatty (Negative)						
Dense (Positive)	72 (TP)	5 (FP)	93.5	84.4	90.82	0.9351	0.8438	0.0917
Fatty (Negative)	5 (FN)	27 (TN)						
Kernel Used: Pearson VII Function-Based Universal Kernel.								
Test Result	Actual		Sensitivity (%)	Specificity (%)	Accuracy (%)	Positive Predictive Power	Negative Predictive Power	Misclassification Rate
	Dense (Positive)	Fatty (Negative)						
Dense (Positive)	75 (TP)	3 (FP)	97.4	90.6	95.41	0.9615	0.9355	0.0459
Fatty (Negative)	2 (FN)	29 (TN)						

It is evident that the results obtained here are in exceptionally good agreement with existing approaches. These results demonstrate that the extracted features have significantly improved the performance of classification. From the experimental results it has been found that, the extracted features are when fed to the SVM classifier with polynomial kernel, gave maximum accuracy of 97.25% with 100% sensitivity and 90.6% specificity. A Sensitivity of 100% means that all positives are being classified as positives i.e. dense mammograms are being recognized as dense that is highly desirable in medical field. With RBF kernel an overall accuracy of 90.82% has been achieved. Therefore this may not be a good option for classification. For Pearson VII function-based universal kernel, an accuracy of 95.41% has been achieved. Sensitivity and specificity are found to be 97.4% and 90.6% respectively.

Experiments demonstrated that the proposed technique gives better results as compared to other approaches suggested by Mustra et al. [25] and Subashini et al. [10] those achieved an accuracy of 91.60% and 95.44% respectively.

This shows that the extracted features are best features in characterizing the texture pattern of fatty and dense mammograms.

CONCLUSIONS

In selecting effective features used in CAD systems for medical images, great research efforts have been focused on identifying and extracting the better features to capture the texture of images and improve correlation to the human visual similarity. In this paper an attempt has been made to classify mammograms on the basis of breast density using texture features. The classification accuracy has been tested on mini-MIAS database. The results provide compelling evidence that the Grey Level Difference Statistics features and Fourier Power Spectrum based features can be used for developing a CAD system for the classification of breast density. An accuracy of 97.25% with 100% sensitivity and 90.6% specificity has been achieved when the extracted features are fed to the SVM classifier with Polynomial Kernel. Thus these models can be explored for disease classification task and retrieval applications in CAD systems.

REFERENCES

1. R.U. Acharya, E. Y. K. Ng, Y. H. Chang, J. Yang and G. J. L. Kaw, “Computer-Based Identification of Breast Cancer Using Digitized Mammograms”, Journal of Medical Systems, Vol. 32, Issue. 6, pp. 499-507, 2008.
2. T.M. Vomweg, “Computer-Aided Diagnosis: Clinical Applications in the Breast”, in Image Processing in Radiology, Springer Berlin Heidelberg, ISBN 978-3-540-25915-2, pp. 355-374, 2008.
3. K. Ganesan, U.R. Acharya, C.K. Chua, L.C. Min, K.T. Abraham and K. Ng, “Computer-Aided Breast Cancer Detection Using Mammograms: A Review”, IEEE Reviews in Biomedical Engineering, Vol. 6, pp. 77-98, 2013.
4. Breast cancer India.net. 2014. [Online]. Available: http://www.breastcancerindia.net. Accessed on 10 Jan, 2014.
5. J.N. Wolfe, “Breast Patterns as an Index of Risk for Developing Breast Cancer”, American Journal of Roentgenology, Vol. 126, Issue 6, pp. 1130-1137, 1976.
6. E.J. Aiello, D.S.M. Buist, E. White and P.L. Porter, “Association between Mammographic Breast Density and Breast Cancer Tumor Characteristics”, Cancer Epidemiology Biomarkers & Prevention, Vol. 14, Issue3, pp. 662-668, 2005.
7. N.F. Boyd, J.M. Rommens, K. Vogt, V. Lee, J.L. Hopper, M.J. Yaffe and A.D. Paterson, “Mammographic breast density as an intermediate phenotype for breast cancer”, The Lancet Oncology, Vol. 6, Issue 10, pp. 798-808, 2005.
8. A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. R. Denton and R. Zwiggelaar, “A Novel Breast Tissue Density Classification Methodology”, IEEE Transactions on Information Technology in Biomedicine, Vol. 12, Issue 1, pp. 55-65, 2008.
9. K. Bovis, and S. Singh, “Classification of mammographic breast density using a combined classifier paradigm” in Medical Image Understanding and Analysis (MIUA) Conference, Portsmouth, pp. 1-4, 2002
10. T.S. Subashini, V. Ramalingam and S. Palanivel, “Automated assessment of breast tissue density in digital mammograms”, Computer Vision and Image Understanding, Vol. 114, Issue 1, pp. 33-43, 2010.

11. S.D. Tzikopoulos, M.E. Mavroforakis, H.V. Georgiou, N. Dimitropoulos and S. Theodoridis, “A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry”, Computer Methods and Programs in Biomedicine, Vol. 102, Issue 1, pp. 47-63, 2011.
12. I.M. Ibrahim and M.A. Wahed, “Visual versus Statistical Features Selection Applied to Mammography Mass Detection”, Journal of Medical Imaging and Health Informatics, Vol. 4, Issue 2, pp. 237-244, 2014.
13. J. Suckling, J. Parker, D. R. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S-L Kok, P. Taylor, D. Beta and J.Savage, “The Mammographic Image Analysis Society Digital Mammogram Database”, in ExerptaMedica. International Congress Series, Vol. 1069, pp. 375-378. 1994.
14. V. Sharma and S. Singh, “CFS-SMO based classification of breast density using multiple texture models”, Medical & Biological Engineering & Computing, Vol. 52, Issue 6, pp. 521-529, 2014.
15. V. Sharma and S. Singh, “Automated Classification of Fatty and Dense Mammograms”, Journal of Medical Imaging and Health Informatics, Vol. 5, Issue 3, pp. 520-526, 2015.
16. H. Li, M. L. Giger, Z. Huo, O.I. Olapade, L. Lan, B. L. Weber and I.Bonta, “Computerized Analysis of Mammographic Parenchymal Patterns for assessing Breast Cancer Risk: Effect of ROI size and Location”, Medical Physics, Vol. 31, Issue 3, pp. 549–555, 2005.
17. J.S. Weszka, C.R. Dyer and A. Rosenfield, “A Comparative Study of Texture Measures for Terrain Classification”, IEEE Transactions on Systems, Man. &Cybernetics,Vol. 6, Issue. 4, pp. 269-285, 1976.
18. G.G. Lendaris and G.L. Stanley, "Diffraction pattern sampling for automatic pattern recognition", in Proceedings of the IEEE, Vol. 58, Issue 2, pp. 198-216, 1970.
19. A.B. Hur and J. Weston, “A user’s guide to support vector machines”, in Data mining techniques for the life sciences, Methods in Molecular Biology, Vol. 609, pp. 223-239, 2010.
20. J.E.E.Oliveira, A.P.B. Lopes, G.Chavez, A.D.A. Araújo, T. M. Deserno, “MammoSVD: a Content-Based Image Retrieval System Using a Reference Database of Mammographies”, 22nd IEEE International Symposium on Computer Based Medical Systems, pp. 1-4, 2009.
21. Y. Yang,L. Wei and R.M. Nishikawa, “Microcalcification Classification Assisted by Content-Based Image Retrieval for Breast Cancer Diagnosis”, Pattern Recognition, Vol. 42, Issue 6, pp. 1126-1132, 2009.
22. C. H. Wei, Y. Li and P.J. Huang, “Mammogram retrieval through Machine Learning within BI-RADS Standards,” Journal of Biomedical Informatics, Vol. 44, Issue 4, pp. 607–614, 2011.
23. T. M. Deserno, M. Soiron, J. E. E. Oliveira and A. D. A. Araujo, “Towards Computer-Aided Diagnostics of Screening Mammography using Content-Based Image Retrieval,” Proceedings of 24th SIBGRAPI Conference on Graphics, Patterns and Images, Washington, USA, pp. 211-219, 2011.
24. S. Ergin and O. Kilinc, “A new feature extraction framework based on wavelets for breast cancer diagnosis”, Computers in Biology and Medicine, Vol.51, pp. 171-182, 2014.
25. M. Mustra, M. Grgic and K. Delac, “Breast Density Classification Using Multiple Feature Selection”,

AUTOMATIKA: Journal for Control, Measurement Electronics, Computer, Vol. 53, Issue 4, pp. 362-372, 2012.

AUTHOR DETAILS

Vipul Sharma is Assistant Professor in Department of Computer Science and Engineering, DAV University, Jalandhar. He has done B.Tech. in Computer Science & Engineering from Punjab Technical University and M.E. in Computer Science & Engineering from Panjab University Chandigarh. He has submitted his Ph.D. thesis in the area of Medical Image Retrieval. He is an active member of IEEE, ACM, and SPIE. His areas of interests include Medical Image Processing, Image Retrieval, and Web Mining.

Lokesh Jain is Assistant Professor in School of Electrical Engineering and Information Technology, College of Agricultural Engineering & Technology, Punjab Agricultural University, Ludhiana, India. He has done B.E in Electronics and Instrumentation, M.Tech in Electrical engineering and M.Tech in Computer Science and Engineering. He has completed his PhD in Faculty of Engineering and Technology. He has overall experience of more than eleven years in teaching and research. He has joined as Scientist in Indian Council of Agricultural Research and has been posted at Central Institute for Research on Cotton Technology, Mumbai in year 2003. In 2005, he joined Punjab Agricultural University, Ludhiana and is serving here till now. His field of interest includes agricultural instrumentation, information and communication systems, mobile and web based application development. He is actively involved in the development and implementation of various software applications for the university